Molecular modelling and Function Prediction of hABH7, human homologue of *E. coli* ALKB7

Shankaracharya*, Das S, Vidyarthi AS

Department of Biotechnology, Birla Institute of Technology, Mesra, Ranchi – 835215,

Jharkhand (India)

E-Mail: - shankaracharya@bitmesra.ac.in , shankaracharya2222@gmail.com

Contact: - +91-651-2276223 (O), +91-493-1978640 (M), +91-651-2275401 (Fax)

* To whom correspondence should be addressed

KEYWORDS: - ALKBH7, ALKB7, Function prediction, hABH7, Molecular modeling, Structure prediction

ABSTRACT

Human homologues of ALKB protein have shown the prime role in DNA damaging drugs, used for cancer therapy. Little is known about structure and function of hABH7, one of the members of this superfamily. Therefore, in present study we are intended to predict its structure and function using various bioinformatics tools. Modeling was done with modeller 9v7 to predict the 3D structure of the hABH7 protein. Tertiary structure model of hABH7, ALKBH7.B99990002.pdb was predicted and evaluated. Validation result showed 97.8% residues in favored and additional allowed region of Ramachandran plot. Ligand binding residues prediction showed four ligand clusters, having 25 ligands in cluster 1. Importantly, presence of Phe120-Gly121-Gly122 conserved pattern in the functional domain was detected. In the predicted structural model of hABH7, amino acid residues, Arginine at 57, 58, 59 and 60 along with tyrosine at 61were predicted in RNA binding sites of the model. The predicted and validated model of human homologue hABH7 resulted from this study may unveil the mechanism of DNA damage repair in human and accelerate the research on designing of appropriate inhibitors aiding in chemotherapy and cancer related diseases.

INTRODUCTION

Escherichia coli AlkBs and their human homologues (hABH) have expanded the concept of alkylation repair by direct reversal method. These are alkylating agents involve in oxidative demethylation of 1-methyladenine and 3-methylcytosine (Mishina et al., 2006) and play an active role in triggering cell's response to DNA damage. Earlier bioinformatics methods had been used to show the relatedness between the different human homologues of Alkb proteins (Kurowski et al., 2003) as well as for the theoretical investigations of structure and function of hABH1 (Shankaracharya et al., 2010a), hABH4 (Shankaracharya et al., 2010b) and hABH5 (Shankaracharya et al., 2011) proteins. It was also observed, in another study, that majority of the bacterial AlkB proteins are DNA repair enzymes, and some of these proteins do not primarily target methylated bases (Born et al., 2009).

Some hABH enzymes have been demonstrated to function as nucleic acid demethylases, catalyzing the oxidative demethylation of 1-methyladenine and 3-methylcytosine in DNA and RNA (Aravind et al., 2001; Ducan et al., 2002; Falnes et al., 2002). Eight human AlkB homologues (ALKBH1-8) have been predicted, of which three (ALKBH1-3) have been shown to

exhibit nucleic acid demethylation activity (Kurowski et al., 2003; Falnes et al., 2004; Westbye et al., 2008). Additionally, a DNA lyase activity has been recently described for ALKBH1 that is Fe(II) and 2-Oxoglutarate (2OG) independent (Muller et al., 2009). Expression of ALKBH8 has been implicated in bladder cancer progression. Recently, a tRNA methyltransferase activity of ALKBH8 has been described and implicated in translational decoding (Fu et al., 2010; Shimada et al., 2009; Songe-Moller et al., 2010). Additionally, hABH5 has shown its activity as a direct transcriptional target of hypoxia inducible factor-1 (HIF-1) and was induced by hypoxia in a range of cell types (Thalhammer et al., 2011).

Need for 3D structure of the AlkB homologues in humans and their structural and functional characterization is significant in recent field of research in cancer medicine and cancer molecular biology. Therefore present study focuses on to the modeling of the 3D structures of hABH7 homologue in humans to understand the characteristic features and to predict its function. Moreover homologues related to cancer therapy if modeled would ease out a way to design inhibitors aiding in chemotherapy.

MATERIALS AND METHODS

Search and retrieval of target protein sequence

Information about protein sequence of human analogue of Alkb (hABH7) was retrieved from NCBI (http://www.ncbi.nlm.nih.gov/).

Selection of template

Template was selected by homology search of query protein (hABH7) sequence against the databases available on PDB (http://ww.rcsb.org) using mGenThreader (Jones, 1999) method. Using mGenThreader web server, templates were selected using fold assessment between target and template.

Homology modeling and evaluation of Model

Homology modeling was done using Modeller 9v7 (Fiser and Sali, 2003; Sali and Blundell, 1993). Difficult modeling was used as the identity between target and template sequences was less. This requires one sequence of known 3D structure and Python 2.5 script files containing Modeller commands. The co-ordinate file of template from PDB was used as such. The predicted model was validated with the program Procheck (Laskoswki et al., 1993) and Ramachandran plot statistics was used to evaluate the stability of the model.

Protein structure accession number

The refined homology model of 3D structure of Habh5 of human was submitted to PMDB (http://mi.caspur.it/PMDB/) (Castrignano et al., 2006) and the same was assigned the identifier PM0076288.

Function prediction

3d2GO server was used for prediction of functions of the predicted model using sequence and structure in the reference of Gene Ontology (GO). It predicts the function of the protein using sources of information like overall topological similarity to structures with known function, geometric and residue similarity of predicted functional sites to regions of known structures and sequence homology to functionally annotated sequences. Then all these information was processed by a Support Vector Machine trained to discriminate between true and false positive functional assignments (http://www.sbg.bio.ic.ac.uk/phyre/pfd/). MAMMOTH structural alignment program was used for full topology search of the model (Ortiz et al., 2002). MUSCLE program was used for functional site prediction of the predicted model (Edgar and Robert, 2004). Functional residue prediction was done using the Jenson-Shannon Divergence (JS Divergence), an information-theory approach to determine relative residue conservation (Capra and Singh, 2007). Such conservation is related to the functional importance of residues. After the finding of the functional site residues, the site was scanned against structures of known function using a fast geometric hashing technique (Moll and Kavraki, 2008).

3DLigandSite prediction

Protein ligand binding residues was predicted using program 3dLigandSite using Critical Assessment of protein Structure Prediction experiment (CASP) (Wass and Stemberg, 2009). This was based on an approach to identify binding sites by combining the use of the predicted structure of the targets with both residue conservation and the location of ligands bound to homologues structures.

RNA binding residue Prediction

RNA interface residue prediction from protein 3D structure was done with **KYG**, a 3D structure based prediction of RNA interface residues in a protein (Kim et al., 2006). It is available at http://cib.cf.ocha.ac.jp/KYG/.

RESULTS AND DISCUSSION

Search for template on National Centre for Biotechnology Information has generated only few homologous structure hits of low identities. Hence difficult modeling method of modeller was used to model the 3D structure of hABH7. Human ABH3 (pdb id 2IUW) was selected as template using mGenThreader tool (Jones, 1999) on the basis of best NetScore (77.740) out of various other related parameters (Table 1).

Table 1: Selection of template from mGenThreader fold recognition search

Conf.	**NetScore**	**p-value**	**PairE**	**SolvE**	**Aln Score**	**Aln Len**	**Str Len**	**Seq Len**	**PDB ID**
CERT	**77.740**	**7e-07**	**-216.3**	**-9.8**	**401.0**	**168**	**204**	**221**	**2IUW**
CERT	70.851	4e-06	-224.0	-6.5	348.0	172	203	221	3I3Q
CERT	67.320	8e-06	-245.5	-9.6	324.0	156	204	221	3BTX

The protein sequences of target (hABH7) and template hABH3 (PDB ID- 2IUW) were aligned and the result is shown in figure 1. The asterisk showed the identity of amino acids present in two protein sequences.

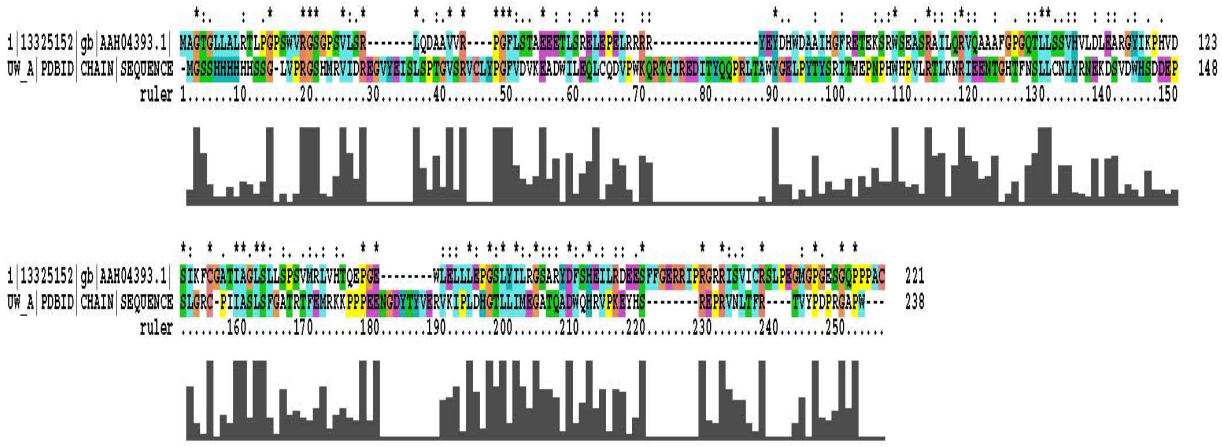

Figure 1: MSA result of hABH7 with the template sequence of 2IUW

Total 5 models were generated after performing homology modeling with modeller 9v7. Dope scores of the generated models were calculated using the command model-single.py. The model ALKBH7.B99990002.pdb, having minimum dope score was considered as the best model of protein hABH7 (Table 2). This result was also supported by the minimum Molpdf scores among five models.

Table 2: Dope energy and related information about successfully produced models

Sl. No.	Filename	Molpdf	DOPE score	GA341 score
1	ALKBH7.B99990001.pdb	1593.13416	-17808.73633	0.29553
2	**ALKBH7.B99990002.pdb**	**1344.40112**	**-17919.59375**	**0.30598**
3	ALKBH7.B99990003.pdb	1439.72168	-17687.18555	0.24658
4	ALKBH7.B99990004.pdb	1462.97925	-17809.68359	0.26922
5	ALKBH7.B99990005.pdb	1696.39453	-17504.89844	0.21872

Further validation program, Procheck (Laskoswki et al., 1993) was used to perform full geometric analysis as well as stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. After running Procheck, Ramachandran plot (Figure 2) shows that for the model hABH5.B99990007, 89.0% residues were in favored region, 8.8% in the additional allowed region, 2.2% in the generously allowed region and none of the residues found in the disallowed region, which made this model more acceptable as compared to other predicted models (Table 3). Homology modeling study is an important method to know the 3D structure of the protein whose structure is not available (Kurowski et al., 2003). Similar approach was also used in the prediction of 3D structure of vaccine related kinaase1 (vrk1) protein (Shankaracharya et al., 2010c), Tubulin β-1 (Shankaracharya et al., 2010d), CDCP2 (Shankaracharya et al., 2010e) and cyclin dependent kinase 4 protein (CDK4) (Shankaracharya et al., 2010f) to predict the respective stable structures and their functionality.

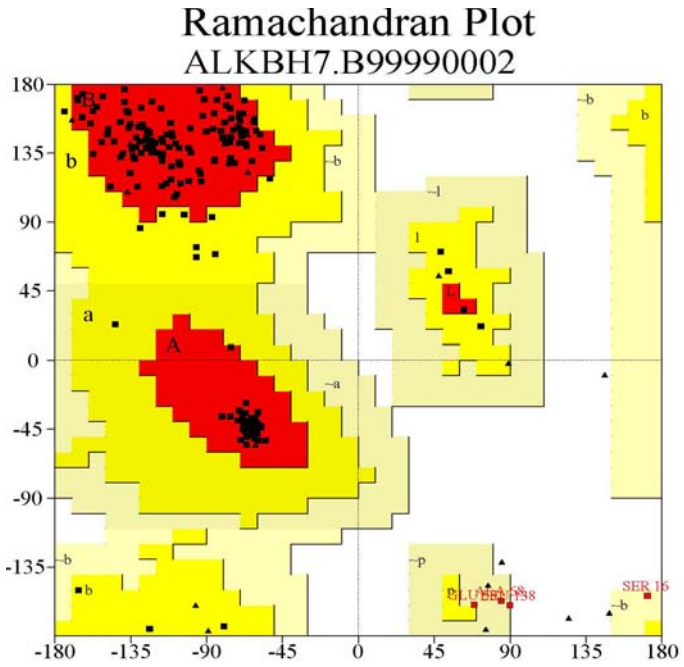

Figure 2: Ramachandran plot of the best model (**ALKBH7.B99990005.pdb**) predicted. Here out of total 221 residues present in the model, 162 lies in most favored region, 16 in additionally allowed region, 4 in generously allowed region and no residues lie in disallowed region.

Table 3: Comparative analysis of Ramachandran Statistics in all the five predicted models

Predicted Structure	Ramachandran Statistics			
	No. of Residues in (%)			
	Most favored Region	**Additional allowed Region**	**Generously allowed Region**	**Disallowed region**
ALKBH7.B99990001.pdb	86.3	9.9	3.3	0.5
ALKBH7.B99990002.pdb	**89.0**	**8.8**	**2.2**	**0.0**
ALKBH7.B99990003.pdb	90.1	8.8	0.5	0.5
ALKBH7.B99990004.pdb	87.4	9.3	2.7	0.5
ALKBH7.B99990005.pdb	84.6	12.1	1.6	1.6

The 3d2GO server was used to predict the function of the protein model. This uses several methods of function prediction, using sequence and structure, to predict Gene Ontology (GO) terms for the protein. Various GO terms, their description and the confidence has been listed in Table 4. Confidence ranges from 0 to 1, with 1 being the most confident prediction. Result show that the predicted protein hABH5.B99990007.pdb has functions like cation binding, ion binding as well as transition metal and metal ion binding with good confidence (Table 4). Two functional sites were also predicted containing amino acid residues as His142, Tyr133, Pro144, Asn131, Iso139 and Pro104 in the first site and Ser59, Iso58, Lys60, Val50, Leu107 and Ala38 in the second. The residues pattern present in the conserved cluster was predicted as Asp56-Phe57-Xn-Pro104-Xn-Asn131-X-Tyr133-Xn-Asp158.

Table 4: Result showing the function prediction of the modeled protein hABH5.B99990005.pdb with 3d2GO (Protein function prediction server)

Sl. No.	GO Term	Description	Confidence
1	GO:0043169	Cation binding	0.92
2	GO:0043167	Ion binding	0.87
3	GO:0046914	Transition metal ion binding	0.85
4	GO:0046872	Metal ion binding	0.85
5	GO:0005488	Binding	0.56
6	GO:0003824	Catalytic activity	0.48

3dLigandSite program was used for the prediction of protein ligand binding residues in Critical Assessment of protein Structure Prediction experiment (CASP). Further the tertiary model of the predicted protein was subjected to the slower but more sensitive structure alignment program MAtching Molecular Models Obtained from Theory (MAMMOTH). The result identified four ligand clusters; among them the first one is most significant predicting 25 ligands as well as 25 structures with average mamooth score of 14.1 (Table 5). In this cluster Phe120, Gly121 and Gly122 residues were predicted in the binding site whose numbers of contacts; average distance and JS divergence have been depicted in Table 6. JS divergence is measured in 0 to 1 scale and higher score mean more conserved residue. Hence the result shows that Phe120, Gly121 and Gly122 are more conserved residue in the structure. In the predicted ligand binding site, heterogens present in the ligand cluster 1 were predicted. The number of each type of ligand and the structures they originated from are also presented (Table 6). Previous study of three-dimensional model prediction for hABH1 active site residues based on other AlkB template 2FD8 has shown that hABH1 contains the five perfectly conserved amino acids in the AlkB family that constitute the iron and 2OG-binding motifs (Westbye et al., 2008).

Table 5: Different ligand clusters information shows that Cluster 1 has maximum numbers of ligands and structures (25 each) with the average Mammoth score of 14.1

Cluster	Ligands	Structures	MAMMOTH Scores		
			Av	Min	max
1	**25**	**25**	**14.1**	**10.6**	**21.5**
2	1	1	11.3	11.3	11.3
3	1	1	15.9	15.9	15.9
4	1	1	15.9	15.9	15.9

Table 6: List of amino acid residues observed in cluster 1 of predicted protein with number of contacts of ligand, Average distance and JS divergence

Residue	Amino acid	Contact	Av distance	JS divergence
120	**PHE**	25	0.00	0.23
121	**GLY**	24	0.04	0.14
122	**GLY**	22	0.47	0.21

Table 6: No. of Counts and list of Heterogens present in the predicted binding site

Heterogen	**Count**	**Source structures**
NI	1	2wwu_A MG 13btx_A
FE	3	2iuw_A,2cgn_A,2cgo_A
FE2	**18**	**2fdj_A, 3i49_A, 3i2o_A, 2fdi_A, 2fdg_A, 2fd8_A,2fdk_A,2g1m_A,1h2n_A, 1mze_A, 2ilm_A,2w0x_A,1mzf_A, 1h2l_A,3hqu_A,2hbt_A,1yci_A,1h2k_A**
ZN	2	3gze_A & 3d8c_A

KYG was used to predict the RNA interface residues on a protein surface (Wass and Sternberg, 2009). The method is based on propensity of residue occurrence in the interface of protein and RNA molecules observed in protein-RNA complex structures. The result shows that residues Arginine at position 57, 58, 59 and 60 along with Tyrosine at 61 are present at the interface of RNA and protein molecule. The similar Structure and function prediction strategies were also used for other human homologue of alkb proteins like hABH1, hABH4 and hABH5. For hABH1, it was found that H231, H287 and D233 were more conserved residue in the structure. The result has also depicted that, residues R24, K25, F27, R28, Y30, R31, Q32, S33, R34, P35 and G36 at the RNA binding site of the predicted protein molecule (Shankaracharya et al., 2010a). However, for hABH4 protein model HIS254, GLU196 and PRO198 were found as more conserved residue in the structure having residues S100, Q101, R104, R105, Q107, D108, Y109, G110, P111, K112, N114, R116, K117, Q118, K119 and K121 at the RNA binding site of the protein molecule (Shankaracharya et al., 2010b). Whereas, in the case of hABH5 protein model, result shows that HIS266, PRO158 and ASP160 are more conserved residue in the structure and residues P354, T355, H356, R357, R358, R359, G360 and S361 are present at the interface of RNA and protein molecule (Shankaracharya et al., 2011).

Therefore, the model developed through homology modeling and subsequently the predicted functional characteristics of hABH7 will initiate the research on identifying a suitable mechanism of repair of alkylation damaged DNA and thus, provide better control on cancer treatment as these DNA repair systems are essential for the maintenance of genome integrity. Consequently, the deregulation of repair genes can be expected to be associated with significant, detrimental health effects, which can include an increased prevalence of birth defects, an enhancement of cancer risk, and an accelerated rate of aging.

CONCLUSION AND PROSPECTS

Homology modeling and function prediction study of hABH7 was performed. The predicted model was validated with program Procheck which shows 97.8% residues in allowed and additionally allowed regions. The ion binding and metal ion binding were predicted as important functional site of the model with high confidence. Amino acid residues pattern of Phe120-Gly121-Gly122 was found as more conserved region in the predicted structure and had been predicted as the most probable ligand binding site in the protein. Further the result also depicted residues Arginine at 57, 58, 59 and 60 as well as tyrosine at 61 are present at the RNA binding site of the protein molecule. These findings are the subject to experimental verification and application for the finding of new chemotherapeutic agent to combat cancer.

ACKNOWLEDGEMENTS

The authors acknowledge BTISnet, Department of Biotechnology, Government of India, New Delhi (No.BT/BI/04/065/04), and the Department of Biotechnology, Birla Institute of Technology, Mesra for providing Infrastructure facility for Bioinformatics Research.

REFERENCES

1. Aravind, L. and Koonin, E. V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent ioxygenases. Genome Biol 2: RESEARCH0007.

2. Born, E. et al. (2009). Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. Nucleic Acids Res. 37(21): 7124 - 7136.

3. Capra, J. and Singh, M. (2007). Predicting functionally important residues from sequence conservation. Bioinformatics. 23:1875.

4. Castrignano, T., et al. (2006). The PMDB Protein Model Database. Nucleic Acids Res. 34(1), D306 - D309.

5. Duncan, T., et al. (2002) Reversal of DNA alkylation damage by two human dioxygenases. Proc Natl Acad Sci. 99: 16660–16665.

6. Edgar, Robert, C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nuc. Acids. Res. 32(5):1792-1797.

7. Falnes, P. O. (2004) Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. Nucleic Acids Res 32: 6260–6267.

8. Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol. 374, 461-469.

9. Fu, D. et al. (2010) Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. Mol Cell Biol 30: 2449–2459.

10. Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J. Mol. Biol. 287, 797-815.

11. Kim, O. T. P., Yura, K., Go, N. (2006) Amino acid residue doublet propensity in the protein-RNA interface and its application to RNA interface prediction. Nuc. Acids. Res. 34 (22), 6450-6460.

12. Kurowski, M. A., Bhagwat, A. S., Papaj, G., Bujnicki, J. M. (2003) Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. BMC Genomics 4: 48.

13. Laskoswki, R. A. et al., (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

14. Mishina, Y., Duguid, E. M., He, C. (2006). Direct Reversal of DNA Alkylation Damage. Chem Rev. 106(2), 215–232.

15. Moll, M. and Kavraki, L. E. (2008) Matching of Structural Motifs Using Hashing on Residue Labels and Geometric Filtering for Protein Function Prediction. The Seventh

Annual International Conference on Computational Systems. Bioinformatics. (CSB2008), Stanford, CA.

16. Muller, T. A., Meek, K., Hausinger, R. P. (2009) Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. DNA Repair (Amst).

17. Ortiz, A. R., Strauss, C. E. Olmea, O. (2002). Mammoth (matching molecular models obtained from theory): An automated method for model comparison. Protein Sci. 11 (11), 2606-2621.

18. Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234, 779-815. Sci U S A 99: 16660–16665.

19. Shankaracharya, Das, S., Vidyarthi A. S. (2010a). Homology Modeling and Function Prediction study of hABH1, Involve in repair of alkylation-damaged DNA, Interdisciplinary Sciences: Computational Life Sciences (Accepted).

20. Shankaracharya, Das, S., Prasad, D., Vidyarthi, A. S. (2010b) Theoretical Investigations on Structure and Function of Human Homologue hABH4 of E.coli ALKB4, *IBC*, 2:8, 1-5 | doi: 10.4051/ibc.2010.2.3.0008.

21. Shankaracharya**,** Srivastava, B., Vidyarthi, A. S. (2010c). Structure modeling of VRK1 Protein and Its Molecular Docking study with Ribavirin analogs. Int. J. Pharm. and Bio Sci. 1(3), pp 1-10.

22. Shankaracharya, Sharma, P., Vidyarthi, A. S. (2010d). Homology Modeling of Tubulin β-1 chain and its Docking study with Colchicine Analogs. Int. J. Pharmacy and Technology. 2(3), 513-527.

23. Shankaracharya**,** Priyamvada, Vidyarthi, A. S. (2010e). Pharmaco-inforamtics: Modeling of Human CDCP2 homologue structure and its docking study with Flavopiridol analogs. Int. J. Pharm. Sci. Rev. Res. 4(1): 1-6.

24. Shankaracharya, Priyamvada, Vidyarthi, A. S. (2010f). Structural Modeling of Human CDK4 and its Docking study with flavopiridol analogues. PharmBIT, 21(1):42-48.

25. Shankaracharya, Das, S., Vidyarthi A. S. (2011) Structure and function prediction of human homologue hABH5 of E. coli ALKB5 using in silico approach, *Nature Precedings*, | doi:10.1038/npre.2011.5597.2.

26. Shimada K., et al. (2009) A novel hhuman AlkB homologue, ALKBH8, contributes to human bladder cancer progression. Cancer Res 69: 3157–3164.

27. Songe-Moller L. et al. (2010) Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. Mol Cell Biol 30: 1814–1827.

28. Thalhammer A. et al. (2011) Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1α (HIF-1α). PLoS One. 6(1):16210.

29. Wass, M. N. and Sternberg, M. J. (2009). Prediction of ligand binding sites using homologueous structures and conservation at CASP8. Proteins. **77** (9), 147-151.

30. Westbye, M. P. et al. (2008). Human AlkB homologue 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. J. Biol. Chem. 283, 25046–25056.